

## MIT Open Access Articles

*Underwater Data Collection Using Robotic Sensor Networks*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Hollinger, Geoffrey A., Sunav Choudhary, Parastoo Qarabaqi, Christopher Murphy, Urbashi Mitra, Gaurav S. Sukhatme, Milica Stojanovic, Hanumant Singh, and Franz Hover. "Underwater Data Collection Using Robotic Sensor Networks." IEEE Journal on Selected Areas in Communications 30, no. 5 (n.d.): 899–911.

**As Published:** <http://dx.doi.org/10.1109/jsac.2012.120606>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <http://hdl.handle.net/1721.1/87723>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Underwater Data Collection Using Robotic Sensor Networks

Geoffrey A. Hollinger, *Member, IEEE*, Sunav Choudhary, *Student Member, IEEE*,  
Parastoo Qarabagi, *Student Member, IEEE*, Christopher Murphy, *Student Member, IEEE*,  
Urbashi Mitra, *Fellow, IEEE*, Gaurav S. Sukhatme, *Fellow, IEEE*, Milica Stojanovic, *Fellow, IEEE*,  
Hanumant Singh, *Member, IEEE*, and Franz Hover, *Member, IEEE*

**Abstract**—We examine the problem of utilizing an autonomous underwater vehicle (AUV) to collect data from an underwater sensor network. The sensors in the network are equipped with acoustic modems that provide noisy, range-limited communication. The AUV must plan a path that maximizes the information collected while minimizing travel time or fuel expenditure. We propose AUV path planning methods that extend algorithms for variants of the Traveling Salesperson Problem (TSP). While executing a path, the AUV can improve performance by communicating with multiple nodes in the network at once. Such multi-node communication requires a scheduling protocol that is robust to channel variations and interference. To this end, we examine two multiple access protocols for the underwater data collection scenario, one based on deterministic access and another based on random access. We compare the proposed algorithms to baseline strategies through simulated experiments that utilize models derived from experimental test data. Our results demonstrate that properly designed communication models and scheduling protocols are essential for choosing the appropriate path planning algorithms for data collection.

**Index Terms**—acoustic communication, path planning algorithms, sensor networks, underwater robotics

## I. INTRODUCTION

THE use of sensor fields to monitor phenomena in underwater environments is of growing interest. Examples include monitoring of algal blooms [1], seismic activity [2], depth surrounding oil platforms, and intrusion of enemy submarines [3]. In underwater scenarios, many standard methods of communication are no longer feasible (e.g., WiFi, cellular, satellite). Acoustic modems can provide communication

G. Hollinger and G. Sukhatme are with the Computer Science Department, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: {gahollin,gaurav}@usc.edu).

S. Choudhary and U. Mitra are with the Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: {sunavcho,ubli}@usc.edu).

P. Qarabagi and M. Stojanovic are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: {qarabagi,millitsa}@ece.neu.edu).

C. Murphy is with the Bluefin Robotics Corporation, Quincy, MA, 02169 USA (email: chrismurf@bluefinrobotics.com).

H. Singh is with the Deep Submergence Laboratory, Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: {cmurphy,hsingh}@whoi.edu).

F. Hover is with the Center for Ocean Engineering, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: hover@mit.edu).

This research has been funded in part by the following grants: ONR N00014-09-1-0700, ONR N00014-07-1-00738, NSF 0831728, NSF CCR-0120778, and NSF CNS-1035866.

Manuscript received July 31, 2011; revised April 28, 2012.



Fig. 1. Representative example of a sensor deployment on the ocean floor to monitor environmental conditions. Such sensors remain in place for many months. Retrieving data from the sensors during deployment is challenging due to limited communication underwater. A mobile AUV equipped with an acoustic modem can act as a data collection device in this scenario by traversing a path that minimizes travel time and maximizes information gathered.

underwater, but they suffer from severe range limitations and channel variations [4].

Without reliable communication, collecting data from underwater sensor networks becomes a challenging problem. A potential solution is the use of a mobile autonomous underwater vehicle (AUV) equipped with an acoustic modem to gather data from the sensors. The problem now becomes one of planning the AUV's path to minimize its travel time and maximize information gathered. We will refer to this problem as the Communication Constrained Data Collection Problem (CC-DCP).

The CC-DCP is closely related to the classical Traveling Salesperson Problem (TSP) [5]. The key difference is that information is gathered from sensors through a noisy channel, the reliability of which decreases with distance and can be modeled probabilistically. Thus, we are dealing with the new problem of *TSP with probabilistic neighborhoods*. While executing the data collection, improved performance can be achieved if the AUV communicates with multiple nodes at once. To this end, we examine canonical versions of two multiple access protocols, one based on Time Division Multiple Access (TDMA) and one on Random Access (RA). We compare these protocols to determine their relative benefits, and we use the results to select parameters for the AUV path planning algorithm.

In this paper, we design path planning algorithms and communication protocols for the application of an AUV gathering

data from an underwater sensor network. The novelties of this paper include: (1) formulation of the Communication Constrained Data Collection Problem (CC-DCP) as a Traveling Salesperson Problem (TSP) with probabilistic neighborhoods, (2) extension of path planning algorithms for the deterministic prize-collecting TSP and TSP with neighborhoods for use in the CC-DCP, (3) comparison of two multiple access protocols for underwater data collection, and (4) the validation of proposed solutions to the CC-DCP through simulated experiments utilizing communication models derived from experimental data. We presented a preliminary version of this work in a prior conference paper [6]. The present paper extends the conference version with data from an AUV deployment, more advanced acoustic communication models, and the comparison of multiple access scheduling protocols.

The remainder of this paper is organized as follows. We first examine related work (Section II) and then formulate the CC-DCP (Section III). We next propose a number of algorithms for solving the CC-DCP approximately (Section IV). We then develop realistic acoustic communication models from experimental test data (Section V), and we derive two multiple access protocols for use during data collection (Section VI). We validate our approach through simulated deployments (Section VII) before concluding and discussing the avenues for future work (Section VIII).

## II. RELATED WORK

The underwater data collection problem is closely related to TSP. In TSP, a mobile agent must visit a number of locations in the minimal amount of time. This problem is known to be NP-hard and has a long history of both approximate and optimal solutions. With current methods, it is possible to solve regular TSP instances optimally with more than 10,000 locations [5]. The two key differences between our work and the classical TSP are: (1) the locations are associated with sensor measurements that may provide different amounts of information regarding the phenomena of interest, and (2) information is gathered from locations based on a probabilistic communication model over an unreliable channel.

The case where locations are associated with different “prizes” has been studied by a number of researchers. The prize-collecting TSP was originally introduced by Balas [7] and has been extended to a number of related variants [8], depending on the type of path required, restrictions on the prizes, and the appearance of new locations during the tour [9]. Goemans and Williamson proposed an approximation algorithm for prize-collecting TSP using a LP primal/dual scheme that achieves a factor of two guarantee [10]. Slightly better approximation guarantees are also possible at the cost of computation and implementation complexity [11].

TSP with deterministic neighborhoods has also been studied. In this formulation, the agent visits a location by moving within a fixed radius of the location. Constant factor approximation algorithms are available for TSP with neighborhoods for both disjoint and overlapping regions of the same size [12] as well as regions of varying sizes [13]. To our knowledge, the combination of TSP with neighborhoods and prize-collecting

TSP has not been studied, and the use of probabilistic communication models to describe the neighborhoods has also not previously been examined.

Related problems have been studied in the context of robotic data mules. Bhadauria and Isler derived approximation algorithms for multiple data mules that must traverse a sensor field and download data [14]. In their work, downloading time is considered as part of the tour, and the communication radii are assumed to be uniform, fixed, and deterministic (i.e., data from a sensor is known to be accessible at a given location). In the present paper, we utilize a probabilistic acoustic communication model that degrades with distance. Such models have been used to optimize control strategies for both underwater station-keeping [15] and underwater search [3].

Vasilescu et al. developed a system of mobile and stationary nodes for underwater data collection based on the use of both optical and acoustic communication [16]. They described the networking architecture and sensor specifications necessary for underwater data collection and presented field experiments using a mobile network. They assume that the AUV communicates with a single node at once, and they do not examine scheduling protocols for multiple access. Their experiments showed the feasibility of utilizing AUVs for underwater data collection, but the authors left open the problem of path planning for the mobile nodes in large networks.

A common assumption in prior work is that the AUV communicates with a single node at a time while executing the tour [6], [16]. For the multiple access case, there has been much work on the design of medium access control methods for underwater acoustic communication networks in both contention-free [17], as well as contention-based [18], [19] categories. The emphasis in these works has been on compensating for the extensive acoustic propagation delay in ad hoc networks. In contrast, our collection of sensor nodes is fixed, and network information is known a priori by the AUV. In fact, it is the AUV that determines which nodes will access the medium via the neighborhood design.

In terms of methods for multiple nodes accessing the network, the classical techniques employed in radio frequency wireless networks [20] include interference avoiding, such as time-division multiple access (TDMA), code-division multiple-access (CDMA), frequency division multiple access (FDMA), and random access methods. FDMA is impractical for underwater channels, due to the severe limits on available bandwidth. CDMA schemes show promise [21]; however there is higher overhead in determining and distributing the spreading codes for a large-scale sensor network. Thus, we focus herein on the comparison of TDMA and a random-access scheme.

## III. PROBLEM SETUP

We are given a pre-deployed network of  $N$  sensors located in  $\mathbb{R}^{dim}$ . For this paper, we limit analysis to  $dim \in \{2, 3\}$ , which yields the 2D and 3D problems respectively. We assume that the location  $x_n \in \mathbb{R}^{dim}$  is given for each sensor  $n \in \mathcal{N}$ , where  $\mathcal{N}$  is the set of deployed sensors. Each sensor  $n$  contains data for retrieval, which we denote as  $Y_n$ . We define the

information quality of the data as  $\mathcal{I}(Y_n)$ , which corresponds to the expected value of information (e.g., information gain in an inference problem [22], or variance reduction in a regression problem [23]).

In the general case, coupling between the sensor measurements can lead to information being subadditive or superadditive. In the context of data collection, we will assume that information is either additive or subadditive (i.e.,  $\mathcal{I}(Y_n, Y_m) \leq \mathcal{I}(Y_n) + \mathcal{I}(Y_m)$  for all  $n \neq m$ ) and that multiple observations of any  $Y_n$  do not provide additional information. Relaxing either of these assumptions leads to interesting extensions (see Section VIII).

The sensors are assumed to have limited capabilities. Each sensor is capable of transmitting data over a noisy channel. A single mobile vehicle has the capability to communicate with the sensors. The location  $x_v \in \mathbb{R}^{dim}$  of the vehicle is controlled and may be subject to constraints, such as obstacles or vehicle kinematics. Based on these constraints, a *traversal cost*  $c(x_1, x_2)$  is defined for all pairs of points  $x_1, x_2 \in \mathbb{R}^{dim}$ . Traversal cost may be defined as the distance, time, energy, or other quantity necessary to move between two points. We assume that the traversal cost obeys the triangle inequality and that the location of the AUV is known. The *communication quality* of a location degrades with distance:  $\mathcal{C}(x_v, x_n) = f(D(x_v, x_n))$ , where  $D(x_v, x_n) = |x_n - x_v|$ , and  $f$  decreases monotonically with distance.

The path planning optimization problem is to generate a path for the vehicle that retrieves data from the sensors and minimizes the traversal cost of the path. For the following derivation,  $P(x_v, x_n)$  represents the probability that data from a sensor at  $x_n$  is received by a vehicle at  $x_v$  (i.e., the probability that the AUV has successfully received data from the sensor). We will set  $\mathcal{C}(x_v, x_n) = P(x_v, x_n)$ , noting that more general communication quality models can be used. If the information quality of sensors is independent, the expected received information quality at point  $x_v$  can be written as

$$R(x_v) = \sum_{n \in \mathcal{N}} P(x_v, x_n) \mathcal{I}(Y_n). \quad (1)$$

With the independence assumption, we can also calculate the expected received information along a path  $\mathcal{P} = [x_v(1), x_v(2), \dots, x_v(T)]$  as

$$R(\mathcal{P}) = \sum_{n \in \mathcal{N}} \left[ 1 - \prod_{t \in \{1, \dots, T\}} [1 - P(x_v(t), x_n)] \right] \mathcal{I}(Y_n). \quad (2)$$

The equations above assume that (1) the information quality of the sensors is independent, (2) the AUV is always attempting to communicate with all sensors, and (3) that the sensors' replies do not cause packet collisions. We relax these three assumptions in Section VI, which requires the development of more sophisticated scheduling protocols, as well as techniques to calculate the information quality  $R(\mathcal{P})$  at a given AUV location. Given an expression for  $R(\mathcal{P})$ , we can state the Communication-Constrained Data Collection Problem (CC-DCP) formally.

**Problem 1:** Given path costs  $c$ , information quality  $\mathcal{I}$ , communication quality  $\mathcal{C}$ , and a set of possible AUV paths  $\Psi$ , find

$$\mathcal{P}^* = \operatorname{argmax}_{\mathcal{P} \in \Psi} R(\mathcal{P}) - l \sum_{t \in \{1, \dots, T\}} c(\mathcal{P}(t-1), \mathcal{P}(t)), \quad (3)$$

where  $T$  is the index of the last location on the path, and  $l$  is a scaling parameter that adjusts the relative tradeoff between information quality and traversal cost.

In some cases, we may want to set the hard constraint to gather information from all sensors. For this case, the optimal solution is a mapping  $2^N \times \mathbb{R}^{dim} \rightarrow \mathbb{R}^{dim}$  from the  $2^N$  possible *received* or *not received* states of the  $N$  sensors and the current location of the AUV to the next best location.

#### IV. ALGORITHMS

We now present algorithms for solving the CC-DCP both optimally, at the cost of high computation, and heuristically, based on existing algorithms for TSP variants.

##### A. Optimal MDP Algorithm

The optimal solution to the CC-DCP can be encoded as a policy mapping from states to actions. To see this, note that the problem can be formulated as a Markov Decision Process (MDP) [24]. In the 3D problem, the states in the MDP are defined as  $\mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times 2^N$ , where  $\mathcal{X}$ ,  $\mathcal{Y}$ , and  $\mathcal{Z}$  are the coordinate spaces for the location of the AUV. The  $2^N$  states represent whether or not data has been collected from each of the  $N$  sensors. Let  $x_s$  be the 2D or 3D AUV location encoded in  $s$ . The reward function  $\mathcal{R}(s, s') = \sum_{n \in \mathcal{F}} \mathcal{I}(Y_n) - l c(x_s, x_{s'})$ , where  $\mathcal{F}$  is the set of sensors that go from uncollected to collected between  $s$  and  $s'$ ,  $c$  is the movement cost, and  $l$  is a scaling parameter. The state transitions  $\mathcal{T}(s, s')$  are defined by the communication model and the motion model of the AUV.

If the environment is discretized, the MDP above can be solved using any standard method (e.g., value iteration or policy iteration) [24]. However, it is important to note that the number of states is exponential in the number of sensors and polynomial in the size of the environment (to the power of 2 or 3 depending on the dimension). Thus, we can expect optimal solutions to become infeasible for any instance of more than a few sensors. In addition, there may be additional errors that arise from discretizing the possible locations of the AUV.

##### B. Approximate Algorithms

Due to the computational intractability of the optimal solution for large networks, we develop heuristics for solving the CC-DCP approximately. The key difference between the CC-DCP and the TSP with neighborhoods [12] is that communication is modeled probabilistically. Our approach is to generate contours of equal probability around the sensors and utilize these as if they were deterministic neighborhoods.

We define a *probabilistic neighborhood*  $\mathcal{G}_n \subset \mathbb{R}^{dim}$  as all locations  $x_v$  where the probability of successful data transfer  $P(x_v, x_n)$  is greater than  $p$ . The value of  $p \in [0, 1]$  determines how conservative the probabilistic neighborhood is. As  $p \rightarrow 1$ ,



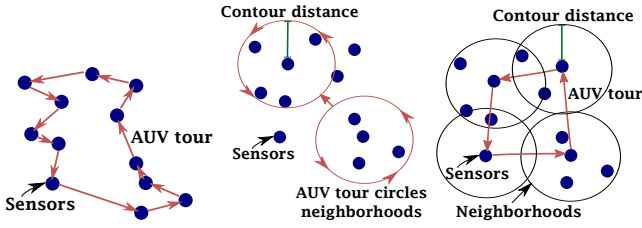


Fig. 2. Example tours using different neighborhood types. **Left:** Standard traveling salesperson tour [5]. **Center:** Tour circling a maximal independent set of neighborhoods [12]. **Right:** Tour visiting the center of a covering set of neighborhoods. All tours travel within the contour distance of all nodes, but the covering set tour is shortest.

it will be near certain that information will be received from sensor  $n$  if the AUV is within the neighborhood. As  $p \rightarrow 0$ , the AUV may need to query a sensor multiple times before receiving data from it.

Once the probabilistic neighborhoods are defined, we can generate a maximal independent set (MIS) of neighborhoods by greedily choosing sensors and removing adjacent sensors with overlapping neighborhoods. The greedy algorithm chooses the sensor neighborhood that covers the largest number of additional sensor neighborhoods and then removes those neighborhoods from further consideration. A valid tour can then be found by circling the neighborhoods in the MIS. We note that the resulting algorithm has a constant factor performance guarantee relative to optimal in the case of deterministic neighborhoods [12].

In the case of probabilistic neighborhoods, it may be necessary to wait for information to be received from the entire neighborhood before moving to the next neighborhood. In addition we make the following modification: instead of generating an MIS and circling the entire neighborhoods, we generate a covering set at half the contour distance. This allows us to visit all sensors by simply planning a tour of the locations in the covering set. This modification improves performance in practice and also allows straightforward extension to 3D environments. In the following sections, we will denote the covering set of neighborhoods as  $\mathcal{G}$ . Figure 2 gives a visualization of planning with varying neighborhood types.

For the case where all nodes must be visited, we design the following heuristics for planning the path of the AUV:

- **Myopic:** Move towards closest sensor. Once data is received, move to next sensor.
- **TSP Solution:** Find an optimal TSP ordering of the sensors using the Concorde solver [5]. Visit sensors in that order. Shortcut sensor once data is received.
- **TSP Solution with neighborhoods:** Find a covering set of probabilistic neighborhoods. Find the optimal TSP ordering of the covering set using Concorde. Visit the neighborhoods in that order. Shortcut once data is received.

The myopic strategy is simply to move towards the closest sensor. This is a reactive strategy and will perform well when communication quality is high. In such cases, the benefit of long-term planning is negated by the homogeneity of the expected received information across the environment. The

second approach is to solve the TSP of the sensors and ignore the communication model. This technique will perform well when communication is poor, since this situation requires the AUV to move near all sensors to gain information from them. Finally, the TSP solution with neighborhoods incorporates the communication model as a probabilistic neighborhood.

For the case where all sensors do not need to be visited, we propose the use of a prize-collecting TSP algorithm (PC-TSP) to improve performance. The PC-TSP assigns a penalty  $\zeta(Y_n)$  to each location based on its information content. The tour now has the option of neglecting some locations and paying the required penalty. The total cost of the tour is then the movement cost plus  $\zeta(n)$  for all  $n$  not visited. In our case,  $\zeta(n) = l \mathcal{I}(Y_n)$ , where  $l$  is a scale factor. We employ the following strategies for the prize-collecting case:

- **Myopic:** Ignore the penalties and act as above. Terminate if the sum of remaining benefits is less than minimum distance to a sensor's probabilistic neighborhood.<sup>1</sup>
- **PC-TSP Solution:** Use the primal/dual algorithm from [10] to determine sensors to visit. Find optimal ordering of this subset using the Concorde solver [5]. Visit sensors in that order. Shortcut once data received.
- **PC-TSP Solution with Neighborhoods:** Find a covering set of the sensors using probabilistic neighborhoods. Use the primal/dual algorithm on the covering set to determine a subset to visit. Find the optimal ordering of the subset using Concorde [5]. Visit in that order. Shortcut once data is received.

The non-myopic algorithms for the case of differing sensor information utilize an existing PC-TSP approximation algorithm to determine which sensors (or neighborhoods) to visit during the tour. The selected locations are then treated in a similar fashion to the problem with equal sensor information. Additional implementation details and derivation of performance guarantees for the prize-collecting case are available in the conference version of this paper [6].

## V. ACOUSTIC COMMUNICATION

We now discuss the acoustic communication model that we utilize to improve path planning for the AUV in the data collection scenario. Acoustic propagation is characterized by energy spreading and absorption that occur in an unobstructed medium over a single propagation path, as well as by additional distortions caused by multipath propagation (i.e., surface-bottom reflections and refraction due to sound speed variation with depth [25]).

While there is no well-accepted acoustic channel model, statistical approaches and geometric approaches are both available. Ray tracing, a geometric approach, offers an accurate picture of the resulting sound field at a given frequency and a given location in a ocean, which can be used to predict signal strength prior to system deployment. However, the actual signal strength, observed in a finite bandwidth and over finite intervals of time, deviates from the predicted value.

<sup>1</sup>An alternative myopic strategy is to move to the node with highest benefit/cost ratio. In practice, this did not perform as well as the simple nearest neighbor.

These variations appear as random, and our goal is to describe them statistically based on experimental data. Such data-driven models allow us to capture environmental factors typical of an AUV deployment.

#### A. Data from AUV Deployment

We utilize data acquired by the AUV *Lucille*, shown in Figure 3. *Lucille*, a SeaBED-class AUV [26] operated by the NOAA Northwest Fisheries Science Center, is equipped with a WHOI Micro-Modem [27] and a 12.5 kHz ITC-3013 hemispherical transducer for acoustic communications. In September of 2010, *Lucille* assisted in mapping the submerged portion of the San Andreas Fault off Northern California, at approximately 39°50'N, 124°W. During this survey, the AUV's onboard networking stack periodically transmitted a three-second long packet. These packets were modulated using Phase Shift Keying (PSK) and transmitted using 4 kHz bandwidth around a center frequency of 10 kHz.

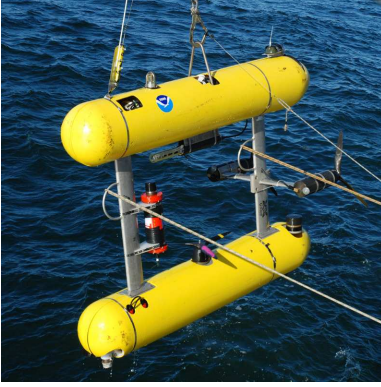


Fig. 3. The AUV *Lucille* being recovered after a successful dive. The ITC transducer used during this experiment is visible protruding upwards from the back of the bottom hull, below the thrusters. Image courtesy of the San Andreas Fault 2010 Expedition, NOAA-OER.

Throughout the course of the dive, the vehicle maintained a constant altitude three meters above the seafloor, at a depth of approximately 130 m. The surface ship, the R/V Pacific Storm, received packets from the AUV at varying slant ranges from 100 m to 1 km. The surface ship remained underway with the hydraulics running during this experiment, resulting in significant noise being generated across the frequencies used for communication. These conditions are typical of those experienced by AUVs operating from near-shore vessels on the continental shelf, and our proposed statistical model incorporates these effects.

#### B. Acoustic Link Gain Model

To specify a propagation model, we represent the link gain as

$$g(d, t) = \bar{g}(d) + y(t), \quad (4)$$

where  $\bar{g}(d)$  is the mean value of the gain at a distance  $d$ ,<sup>2</sup> and  $y(t)$  is a random process.

<sup>2</sup>The distance is varying with time, i.e.  $d = d(t)$ .

We now proceed to establish two models based on our experimental data: one that relates the mean value  $\bar{g}$  to the distance  $d$ , and another that specifies the probability distribution function (pdf) of the random component  $y$ . These models will be valid for the chosen operating conditions (frequency band and transmission distances). Specifically, we make the following conjectures:

(i) the mean value obeys a log-distance model

$$\bar{g}(d) = g_0 - k_0 \cdot 10 \log d, \quad (5)$$

(ii) the random component obeys a Gaussian distribution,  $y \sim \mathcal{N}(0, \sigma^2)$ .

Figure 4 summarizes the recorded values (from the deployment described above) of the gain as a function of distance. The solid curve represents the log-distance model (5), whose parameters  $g_0$  and  $k_0$  were obtained by first-order polynomial fitting.<sup>3</sup> We emphasize again that the model parameters will in general depend on the operational conditions [28], i.e. that the values indicated in the figure are representative of the 8-12 kHz acoustic band and transmission distances on the order of several hundreds of meters.

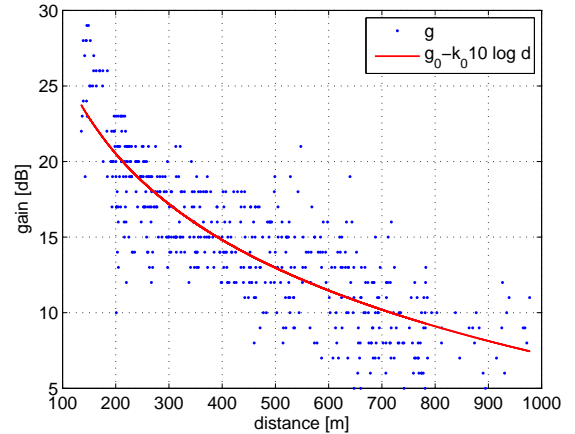


Fig. 4. Gain (normalized) vs. transmission distance. Dots show measured values; solid curve shows an estimated trend (a first-order logarithmic-scale polynomial fit to the ensemble mean at each distance yields  $k_0=1.9$ ).

Shown in Figure 5 is the histogram of the random component  $y = g - \bar{g}$ . This figure motivates our second conjecture, i.e. the Gaussian model for  $y$ . The variance  $\sigma^2$  is calculated from the data at hand. We note that its value appears to be invariant for the range of distances considered, although greater distance spans could require sectioning. We also note that the variance will depend on the bandwidth, decreasing as the bandwidth increases. Similar conclusions have been found using different data sets [29].

#### C. Packet Error Approximations

In addition to the log-normal link gain model justified in the prequel, we adopt a previously employed [3], [15] colored noise model that incorporates multiple environmental factors, including wind, shipping activity, thermal noise, and

<sup>3</sup>Logarithms are taken with base 10.

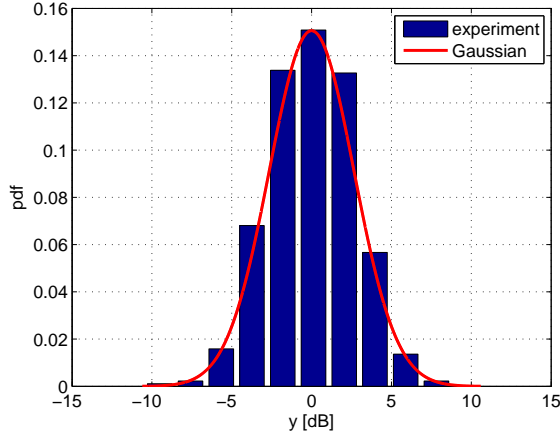


Fig. 5. Histogram of the measured deviation  $y$  and the theoretical p.d.f. of a zero-mean Gaussian random variable with  $\sigma^2=6.7$  dB.

turbulence [25]. We assume a block log-normal fading model for SNR with instantaneous SNR being constant over the duration of one block. We also assume that successive blocks fade independently. With BPSK modulation, the probability of bit error at an SNR of  $\gamma$  is given by:

$$P_b(\gamma) = Q\left(\sqrt{2\gamma}\right), \quad (6)$$

where  $Q(\cdot)$  denotes the tail probability function for the standard normal distribution. For simplicity, we suppress the dependence of  $\gamma$  on transmission range, i.e.  $\gamma = \gamma(d)$ . Assuming  $J$  bits per block, the probability of block error at an SNR of  $\gamma$  is given by:

$$P_S(\gamma) = 1 - (1 - P_b(\gamma))^J \approx J \cdot Q\left(\sqrt{2\gamma}\right), \quad (7)$$

where the approximation holds for large  $\gamma$ . For a packet spanning  $V$  blocks encoded with a code of rate  $r$  to exploit selection diversity, the packet error rate averaged over  $\gamma$  is given by:

$$\overline{P_D} = 1 - (1 - \overline{P_S})^{rV} \approx r \cdot V \cdot \overline{P_S}, \quad (8)$$

where  $\overline{P_S} = \mathbb{E}_\gamma[P_S(\gamma)]$ , and the approximation holds for small values of  $\overline{P_S}$ .

There is no known simple closed form approximation to  $\overline{P_S}$  when  $\gamma$  is log-normally distributed, and so we employ semi-analytic Monte-Carlo methods to compute the packet error. In this model, the packet success rate of  $1 - \overline{P_D}$  between a vehicle at  $x_v$  and a sensor at  $x_n$  represents the expected communication quality  $\mathcal{C}(x_v, x_n)$  (see Section III).

## VI. SCHEDULING PROTOCOLS

We now examine two multiple access protocols to allow the AUV to communicate with multiple nodes at once while executing the tour. We assume that a single carrier, half-duplex, narrowband communication system is present on the AUV as well as on each node. The protocols are described below followed by a performance analysis in the following subsection.

### A. Protocol Description

1) *Time Division Multiple Access (TDMA)*: We assume fixed locations for sensor nodes in each neighborhood and that synchronization (see e.g. [30]) among them has been accomplished. The number of transmission slots per neighborhood is fixed at  $N_s$  for the given tour and length of each slot is sufficient to receive all packets from a single node. A three phase TDMA protocol is described below:

- 1) *Initiation*: The nodes begin in a low power state wherein they can hear broadcast wake-packets from the AUV if in range. The AUV sends a high power broadcast wake-packet of size  $B_S$  that brings into an active state, the nodes within the AUV's current neighborhood (see e.g. [31]). The broadcast wake-packet also contains initial communication schedules for all nodes in the neighborhood.
- 2) *Scheduling*: The functional nodes that receive the broadcast correctly reply with an acknowledgement packet of size  $B_A$  according to the schedule. The AUV sends out the next round of scheduling information to these nodes.<sup>4</sup>
- 3) *Data Transfer*: The nodes reply with data packets. After all nodes have completed their transmissions, this phase restarts for a second round of transmissions under the same schedule. The number of transmission rounds executed is a design parameter.

2) *Random Access (RA)*: We assume that nodes lack carrier-sense capability. We also suppose that any packet collision at the AUV leads to reception error. A two phase unslotted RA protocol is described below:

- 1) *Initiation*: The nodes begin in a sleep state. The AUV sends a high power broadcast wake-packet which brings the nodes into an active state.
- 2) *Data Transfer*: Each node transmits data packets with a random backoff between successive packets. When a node completes its transmission, it restarts for a second round of transmission. The number of transmission rounds executed depends on the neighborhood.

The absence of node scheduling and selection reduces data transfer overhead for RA as compared to TDMA, but a non-zero probability of packet collision decreases throughput. We note that the TDMA and RA protocols do not use acknowledgements (ACKs) for data packets, which simplifies implementation at the expense of gain from feedback and reduces the round trip propagation delay.<sup>5</sup>

### B. Protocol Performance Analysis

We now develop expressions that will allow us to evaluate the performance of the two representative protocols described above. We assume that the AUV receives a packet correctly if at least one of the many transmitted copies of the packet is

<sup>4</sup>A subset of nodes can be easily accommodated as well if AUV constraints such as power necessitate communication with a smaller set.

<sup>5</sup>We consider uncoded performance in Section VII to focus on the interplay between the scheduling protocols and the path planning algorithms. However, error correction can easily be incorporated by increasing the effective channel SNR appropriately.



received correctly. While the transmissions from the nodes to the AUV will incur errors, we assume that transmissions from the AUV to the nodes are perfectly decoded.

Let  $\mathcal{N}$  denote the set of all deployed sensor nodes, and let  $\mathcal{M}$  denote the subset of nodes covered by the AUV in the course of its entire tour. We denote by  $\mathcal{G}$  the set of neighborhoods that are visited by the AUV (see Section IV). The  $|\mathcal{M}|$  sensors are divided into  $|\mathcal{G}|$  neighborhoods, and we assume that each sensor in  $\mathcal{M}$  belongs to exactly one neighborhood in  $\mathcal{G}$ . We define  $h: \mathcal{M} \rightarrow \mathcal{G}$  and  $K: \mathcal{G} \rightarrow \mathbb{Z}^+$  to be functions respectively mapping a node  $m \in \mathcal{M}$  to its neighborhood  $h(m) \in \mathcal{G}$ , and a neighborhood  $b \in \mathcal{G}$  to the number of transmission rounds executed  $K(b) \in \mathbb{Z}^+$ . We note that definition of a transmission round is protocol dependent as described in section VI-A. Without loss of generality, if  $x_v$  is the location of the AUV when at the center of the neighborhood, we can assume the nodes to satisfy  $D(x_v, x_m) \leq D(x_v, x_n)$  whenever node indices  $m$  and  $n$  satisfy  $m < n$  and  $h(m) = h(n)$ . Let  $C_1$  and  $C_2$  denote proportionality constants, respectively transforming physical distance into propagation delay, and packet size into transmission time.

1) *Expected Information Gain*: The expression for expected information gain can be developed independently of the underlying protocol. Let  $Z_i$  denote the data stream of sensor  $i \in \mathcal{M}$  and let  $Z_S$  denote the set of data streams of all sensors in the subset  $S \subseteq \mathcal{N}$ . For now, we assume that all sensors are equally informative meaning that  $H(Z_i)$  is independent of  $i$ , where  $H(\cdot)$  is the entropy function. In an actual deployment, information from different sensors are typically correlated. To capture this effect, we define the *conditional common information* for the sensor pair  $(i, j) \in \mathcal{N} \times \mathcal{N}$  as the quantity  $I(Z_i; Z_j | Z_{\mathcal{N}-\{i,j\}})$  where  $I(\cdot; \cdot | \cdot)$  is the conditional mutual information function. For consistency, we also define the *conditional innovation* for sensor  $i \in \mathcal{N}$  as the quantity  $I(Z_i; Z_i | Z_{\mathcal{N}-\{i\}}) = H(Z_i | Z_{\mathcal{N}-\{i\}})$ . To simplify book keeping, we further suppose that any correlated information in  $\mathcal{N}$  can be captured by the conditional common information for some subset of sensor pairs  $(i, j) \in \mathcal{N} \times \mathcal{N}$ , which essentially means the following is assumed:

$$I(Z_i; Z_j | Z_k) = I(Z_i; Z_j) \quad \forall k \in \mathcal{N} - \{i, j\}. \quad (9)$$

We consider a fixed packet size of  $B_D$  bits and assume that all packets from a single node in one transmission round carry independent but equal quantities of information. This means that we can assume equal number of packets at each sensor in accordance with nodes being equally informative. Let this number be  $N_p$ . We further assume that the AUV performs joint decoding of information streams from different nodes and can distinguish between packets containing common information and innovation packets.

We can now compute the expected information gain at the AUV for an entire tour  $\mathcal{P}$  in terms of the number of correctly received packets carrying distinct information. A packet at sensor  $m \in \mathcal{M}$  is transmitted  $K(h(m))$  times over the node  $m$  to AUV link. So, the probability of packet failure over this

link is given by:

$$\Pr_m(\text{failure}) = \overline{P}_D^{K(h(m))}, \quad (10)$$

where  $\overline{P}_D$  is given by (8) after substituting fading statistics for node  $m$  to AUV link. For  $S \subseteq \mathcal{N}$  with  $|S| = 2$ , let  $N_p(S)$  denote the conditional common information between sensors in  $S$ . Since conditional common information resides at both sensors in  $S$ , the probability of failure of such a packet is given by:

$$\Pr_S(\text{failure}) = \prod_{n \in S} \Pr_n(\text{failure}) = \prod_{n \in S} \overline{P}_D^{K(h(n))}. \quad (11)$$

Using (11) and letting  $N_p(S)$  denote conditional innovation for  $|S| = 1$ , the total expected information gain for an entire tour  $\mathcal{P}$  is given by:

$$\begin{aligned} R(\mathcal{P}) &= \sum_{S \subseteq \mathcal{N}: |S| \leq 2} N_p(S) \cdot (1 - \Pr_S(\text{failure})) \\ &= \sum_{S \subseteq \mathcal{N}: |S| \leq 2} N_p(S) \cdot \left( 1 - \prod_{n \in S} \overline{P}_D^{K(h(n))} \right), \end{aligned} \quad (12)$$

where we assume  $K(h(n)) = 0$  for any sensor  $n \in \mathcal{N} - \mathcal{M}$ .

Equation (12) is directly applicable to TDMA. However, for RA, the packet collision probability needs to be accounted for in the packet error probability. We define  $\lambda: \mathcal{G} \rightarrow \mathbb{R}^+$  to be a function mapping a neighborhood  $b \in \mathcal{G}$  to the average packet arrival rate at the AUV  $\lambda(b) \in \mathbb{R}^+$  per  $C_2 \cdot B_D$  unit of time. Assuming no synchronization between nodes, the number of packets arriving at the AUV at any given time is well approximated by a Poisson process with parameter  $\frac{\lambda(b)}{C_2 \cdot B_D}$  [32]. Let us denote the number of nodes in neighborhood  $b \in \mathcal{G}$  as  $s(b) \in \mathbb{Z}^+$ . Given that a packet arrives at time  $t$  from node  $n$ , a collision occurs if additional packets arrive in the interval  $(t - C_2 \cdot B_D, t + C_2 \cdot B_D)$  and so the probability that no collision occurs is given by:

$$\Pr_b(\text{no collision}) = \exp \left( -2 \cdot \lambda(b) \cdot \left( 1 - \frac{1}{s(b)} \right) \right) \quad (13)$$

$$\approx \exp(-2 \cdot \lambda(b)), \quad (14)$$

where the packet arrival rate  $\lambda(b)$  has been adjusted to exclude node  $n$  and the approximation holds for large neighborhoods. Probability of packet error given that no collision occurs is same as  $\overline{P}_D$ . So, effective average packet error rate for RA is given by:

$$\begin{aligned} \overline{P}_{D\text{eff}} &= \Pr_{h(n)}(\text{collision}) + \Pr_{h(n)}(\text{no collision}) \cdot \overline{P}_D \\ &= 1 - \exp \left( -2 \cdot \lambda(h(n)) \cdot \left( 1 - \frac{1}{s(h(n))} \right) \right) \cdot (1 - \overline{P}_D) \end{aligned} \quad (15)$$

$$\approx 1 - \exp(-2 \cdot \lambda(h(n))) \cdot (1 - \overline{P}_D). \quad (16)$$

We substitute  $\overline{P}_{D\text{eff}}$  for  $\overline{P}_D$  in (12) which gives the total expected information gain for RA.

Although the expression for expected information gain is valid for any information correlation model, for simulation



purposes we consider the following model. We define the *information correlation* between sensors  $m, n \in \mathcal{N}$  as:

$$\rho_{m,n} = \frac{I(Z_m; Z_n | Z_{\mathcal{N}-\{m,n\}})}{H(Z_m)}, \quad (17)$$

and assume it to decay exponentially<sup>6</sup> with separation for  $m \neq n$  as:

$$\rho_{m,n} = \exp\left(-\frac{D(x_m, x_n)}{\alpha}\right), \quad (18)$$

where  $\alpha$  is scaling parameter.

Using (17) and then (9) we get  $\rho_{m,m}$  as:

$$\begin{aligned} \rho_{m,m} &= \sum_{n \in \mathcal{N}} \rho_{m,n} - \sum_{\substack{n \in \mathcal{N} \\ n \neq m}} \rho_{m,n} \\ &= \sum_{n \in \mathcal{N}} \frac{I(Z_m; Z_n | Z_{\mathcal{N}-\{m,n\}})}{H(Z_m)} - \sum_{\substack{n \in \mathcal{N} \\ n \neq m}} \rho_{m,n} \\ &= 1 - \sum_{\substack{n \in \mathcal{N} \\ n \neq m}} \rho_{m,n} \end{aligned} \quad (19)$$

In terms of packets, this means that data streams  $Z_m$  and  $Z_n$  share  $N_p \cdot \rho_{m,n}$  packets of information for  $m \neq n$ .

2) *Expected Communication Cost*: We define  $G: \mathcal{G} \rightarrow \mathbb{R}^+$  to be the function mapping a neighborhood  $b \in \mathcal{G}$  to the cost of communication in that neighborhood  $G(b) \in \mathbb{R}^+$  during the tour. Let  $e(b)$  denote the farthest sensor from the AUV in neighborhood  $b \in \mathcal{G}$ . The total communication cost for the entire tour is then  $\sum_{b \in \mathcal{G}} G(b)$ . The value of  $G(b)$  depends on the communication protocol and is computed below for TDMA as well as RA.

a) *TDMA*: The initiation phase has a broadcast of size  $B_S$ . This must reach the farthest sensor, so round-trip propagation delay is  $C_1 \cdot 2 \cdot D(x_v, x_{e(b)})$ , and the transmission cost is  $C_2 \cdot B_S$ . In the scheduling phase, the reception time for all ACK packets is  $C_2 \cdot B_A \cdot s(b)$  in addition to scheduling and propagation delay equivalent to initiation phase. If  $\tau_{max}$  is the maximum delay spread, we need a guard interval of  $N_s \cdot \tau_{max}$  for each transmission round and  $2 \cdot s(b) \cdot \tau_{max}$  for the initiation phase. Data transfer spans  $K(b)$  transmission rounds with  $N_s$  slots per round,  $N_p$  packets per slot and  $C_2 \cdot B_D$  transmission time per packet. The communication cost for TDMA is given by:

$$\begin{aligned} G(b) &= \text{Initiation Cost} + \text{Scheduling Cost} \\ &\quad + \text{Guard Interval} + \text{Data Transfer Cost} \\ &= 2 \cdot (2 \cdot C_1 \cdot D(x_v, x_{e(b)}) + C_2 \cdot B_S) + C_2 \cdot B_A \cdot s(b) \\ &\quad + 2 \cdot s(b) \cdot \tau_{max} + N_s \cdot \tau_{max} \cdot K(b) \\ &\quad + C_2 \cdot B_D \cdot N_p \cdot N_s \cdot K(b) \\ &\approx C_2 \cdot B_D \cdot N_p \cdot N_s \cdot K(b), \end{aligned} \quad (20)$$

where the approximation holds for long data streams.

b) *RA*: The initiation phase has a round trip propagation delay of  $C_1 \cdot 2 \cdot D(x_v, x_{e(b)})$  like TDMA. Data transfer spans  $K(b)$  transmission rounds with  $N_p$  packets per sensor at a transmission rate of  $\frac{\lambda(b)}{s(b)}$  packets per sensor per  $C_2 \cdot B_D$  unit of time. This gives the communication cost as:

$$G(b) = 2 \cdot C_1 \cdot D(x_v, x_{e(b)}) + \frac{C_2 \cdot B_D \cdot N_p \cdot K(b) \cdot s(b)}{\lambda(b)} \quad (22)$$

$$\approx \frac{C_2 \cdot B_D \cdot N_p \cdot K(b) \cdot s(b)}{\lambda(b)}, \quad (23)$$

where the approximation holds for long data streams.

The expressions derived above provide the basis for simulations that compare the respective benefits of TDMA and random access in the underwater data collection domain. We next employ these expressions to evaluate the performance of the path planning algorithms on a number of simulated deployments.

## VII. SIMULATIONS

A simulation environment was implemented in C++ running on Ubuntu Linux to test the CC-DCP algorithms. The simulated experiments were run on a 3.2 GHz Intel i7 processor with 9 GB of RAM. We first test the AUV path planning component of the proposed algorithm, and we then compare the two multiple access protocols with different parameter settings to optimize communication scheduling during the data collection.

### A. Path Planning Comparison

This section validates the path planning component of the proposed algorithm. The simulations utilize random deployments of 100 sensors in 100 km  $\times$  100 km 2D environments. The simulated AUV moves at 5 km/hr. Random deployments are used to determine average-case behavior of the proposed algorithms. The simulations utilize the model built from the AUV deployment data (see Section V-A). Further experimental validation of the models is an important avenue for future work. We note that the proposed algorithms are general enough to be used with a wide range of communication models that provide a probability of receiving data from a given node that degrades with distance.

For the initial simulations, the node utilities are set to uniform, and the requirement is to visit all nodes. In these large environments, the communication time is negligible when compared to the travel time, so we do not consider the choice of scheduling protocol. The optimal MDP solution using value iteration was able to solve problems with up to 3 nodes on a 15 km  $\times$  15 km environment with a 1 km grid resolution. For these small problem sizes, the myopic nearest neighbor heuristic performed competitively with the optimal solution. Based on this finding, we scale up the size of the instances to compare the approximate methods, and we remove the infeasible optimal method from consideration.

We next compare the myopic nearest neighbor strategy to the TSP strategy and the TSP with probabilistic neighborhoods

<sup>6</sup>The exponential decay model is commonly used in Kriging and Gaussian Process models of spatial correlations [33].

of  $p = 0.8$ . This value of  $p$  was selected empirically, and we give a more thorough analysis of selection of neighborhood size in Section VII-B. Figure 6 shows results from these trials. At low transmission power (poor communication), the value of utilizing the neighborhoods is minimal, and the problem reduces to the classical TSP. At high transmission power (good communication), the value of planning is reduced, and the simple myopic strategy moves closer to the quality of the non-myopic strategies. At moderate communication levels, there is some improvement from both solving the TSP and utilizing neighborhoods. We also compare to a standard lawnmower pattern that determines the coarsest layout of vertical traversals that collect data from all nodes. The lawnmower patterns perform poorly relative to the proposed algorithms because they do not consider the communication neighborhoods or the exact locations of the nodes.

The same simulations were run with random information values from 0 to 25 added to the 100 sensors. The total cost is calculated by summing traveled distance plus the scaled information values of sensors not collected. The scale factor was set as  $l = 1$  for these simulations. Modifying  $l$  would allow for tuning the number of sensors visited. The PC-TSP approximation algorithm was compared with and without neighborhoods to the nearest neighbor strategy and the lawnmower patterns.

Figure 7 shows the numerical results for the prize-collecting case. Solving the underlying PC-TSP without neighborhoods does not perform well, even when compared to the myopic strategy. Since the PC-TSP algorithm cannot account for the cost of the neighborhoods when determining which sensors to visit, it chooses to ignore a number of sensors that would actually improve the final cost. In contrast, the combination of the neighborhoods and the underlying PC-TSP approximation algorithms performs well. Combining Figures 6 and 7, we see that considering neighborhoods helps marginally in the case without prizes and more significantly when prizes are considered. The lawnmower pattern can also be applied to the prize-collecting case; however, in this case the coarsest layout may not yield the lowest cost. We choose the lawnmower pattern that provides the lowest cost, though it may pay the penalty for missing some nodes. Even with these modifications, the lawnmower pattern still performs poorly.

The running time of the algorithms is dominated by the cost of calculating the TSP tour with the Concorde solver. In the worst-case, this computation time can grow exponentially in the number of nodes. In practice, typical instances of 100 nodes were solved in 10-100 ms. For a more extensive discussion of the running time of Concorde, see [5]. When probabilistic neighborhoods are taken into account, the TSP solver uses a reduced set of nodes (those in the covering set). Thus, utilizing neighborhoods actually reduces the running time of the algorithm.

### B. Multiple Access Protocol Comparison

We now examine the performance of the proposed scheduling protocols integrated into the contour-based TSP path planning algorithm. In the following simulations, we allow

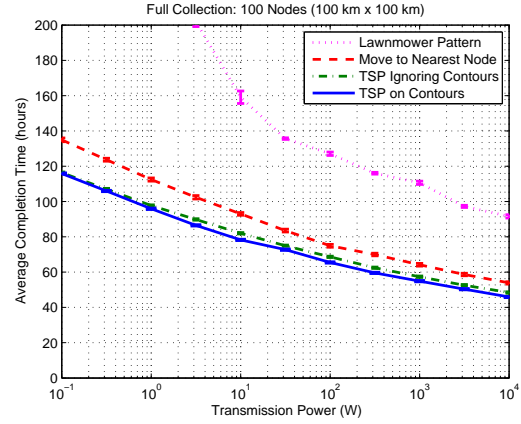


Fig. 6. Communication-constrained data collection simulations. Error bars are one SEM, and averages are over 100 random deployments. Solving the TSP provides improvement over myopic and lawnmower techniques, and utilizing neighborhoods provides some additional improvement.

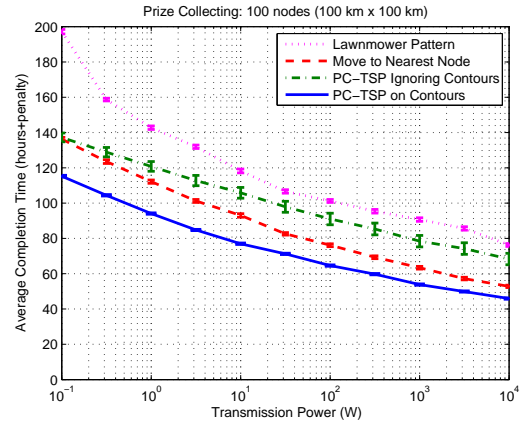


Fig. 7. Prize-collecting communication-constrained data collection simulations. Data are not collected from some sensors, and the required penalty is paid. Error bars are one SEM, and averages are over 100 random deployments. Considering neighborhoods significantly improves performance.

the vehicle to access all nodes in the neighborhood when it reaches the center of the neighborhood. We also assume that no communication occurs while the vehicle is moving between neighborhoods. Relaxing this assumption is an avenue for future work.

A random 2D deployment of 100 sensors was generated in a  $5 \text{ km} \times 5 \text{ km}$  area, and a simulated AUV was added to the environment that moves at a speed of 1 m/s. The size of the environment is smaller than in the previous simulations to explore cases where AUV travel time does not dominate the mission time. The AUV modem was assumed to operate at low power in this smaller environment, which creates a tradeoff between staying in the current neighborhood for additional transmission rounds or moving to the next neighborhood. We considered 200 data packets per node with packet transmission period of 10 ms. A carrier frequency of 13 kHz was used with a maximum channel delay spread of 30 ms. The TDMA protocol was set to use 10 transmission slots per transmission round while the RA protocol was set for each node to transmit 0.2 packets per transmission period. The presented results are

averaged over  $10^5$  Monte-Carlo runs.

Neighborhoods were generated as in the previous simulations, and a TSP tour was calculated to visit all neighborhoods. To focus on the evaluation of the scheduling protocols, we do not consider tours that avoid neighborhoods. Simulations were run for the two multiple access protocols with varying values of the number of transmission rounds and the parameter  $p$ . The value of  $p$  describes the size of the probabilistic neighborhoods, and we note that a higher  $p$  corresponds to smaller neighborhood size (see Section IV). These two parameters represent design decisions when implementing the contour-based TSP algorithm. Figure 8 shows the results of these simulations. As expected, both information gain in Figure 8(a) and communication cost in Figure 8(b) increase as the number of transmission rounds are increased. In addition, increases in  $p$  (corresponding to decreased neighborhood size) lead to longer paths for the AUV and increased cost.

More interesting observations arise when we examine the gain to cost ratio in Figure 8(c). We see that the gain to cost ratio first increases with increasing transmission rounds and then decreases. The maximum appears at a different number of transmission rounds for varying probabilistic neighborhood size. If we examine the effect of correlated information on the choice of scheduling protocol parameters in Figure 8(d), we see that increases in information correlation favor fewer transmission rounds to maximize gain/cost ratio. This result is expected, since correlated data allow for high information gain without receiving data from the entire neighborhood. We observe a different effect when the variance of the link gain  $\sigma$  is increased from Figure 8(e) to Figure 8(f). At higher variances, larger neighborhood sizes and additional transmissions are favored to maximize the probability of receiving data. The ability to handle these various cases by changing the neighborhood size demonstrates the flexibility of the contour-based TSP path planning algorithm in this domain.

If we look at the gain/cost frontier, we see that the solutions can be tuned based on different weightings of cost and gain (see Figure 9). By varying the value of  $p$  and the transmission rounds, we create a frontier of solutions that provides a tradeoff between mission time and information gain.

We note that the TDMA protocol outperforms the RA protocol in terms of gain/cost ratio. In most underwater data collection scenarios, the quantity of information at each sensor is large, which makes TDMA the favored protocol. In applications where the information content at each node is small, poor synchronization and poor communication quality can lead to significant overhead for TDMA. In such cases, RA becomes a better choice for multiple access. In addition, random access allows for ad hoc networking and is easily extended to the case where there is error in the node locations. We note that in many cases, the benefit of TDMA over random access is quite small, and random access may be preferred due to its relative ease of implementation.

## VIII. CONCLUSIONS AND FUTURE WORK

This paper has shown that communication-constrained data collection is feasible with robotic sensor networks. When

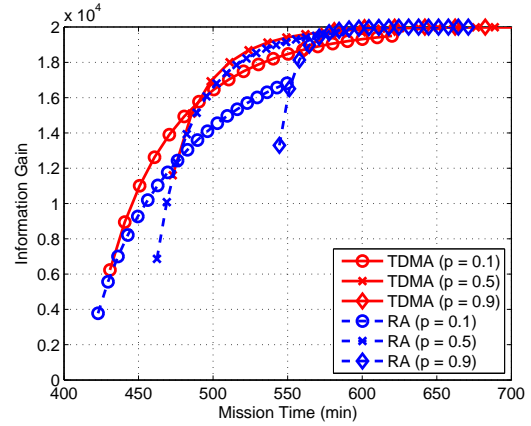


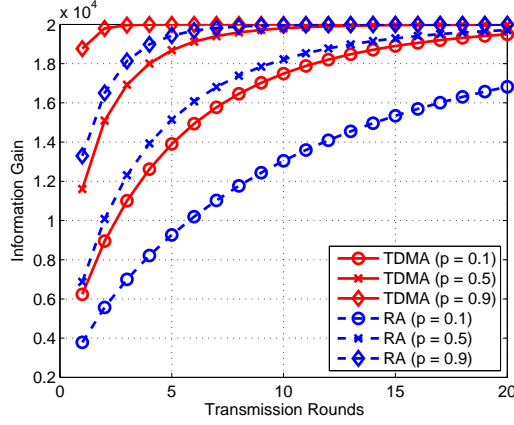
Fig. 9. Communication-constrained data collection simulations in a  $5 \text{ km} \times 5 \text{ km}$  area with 100 nodes. A frontier of solutions is generated that provides a tradeoff between information gain and mission time.

communication quality is high (relative to the sparsity of the network), planning is less essential. When communication quality is low, it is important to solve the underlying TSP, rather than using a fully reactive approach. Finally, at moderate levels of communication, it is beneficial to consider the communication neighborhoods during planning. In addition, when information quality of the sensors is considered and all sensors do not need to be visited, considering neighborhoods improves performance. This analysis provides insight into the level of planning required to optimize information gathering at different levels of network sparsity, communication, and information quality. Such insight motivates the use of realistic communication models in the development and analysis of planning algorithms.

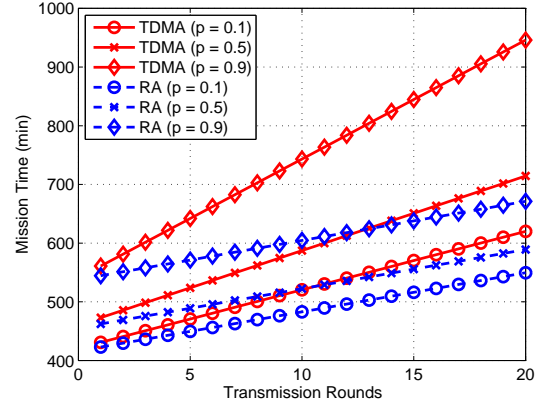
We have also demonstrated the benefit of utilizing scheduling protocols to design path planning algorithms for autonomous data collection. We have shown that simulated analysis with varying parameters can be used to build up a frontier of solutions that tradeoff between mission time and information gain. Without such analysis, it would not be possible to generate this frontier of solutions, and the path planning algorithm would need to execute blindly. Thus, improved scheduling protocols and analysis of communication provide powerful tools for optimizing path planning algorithms in data collection scenarios.

A number of interesting extensions provide avenues for future research. The case where sensor communication quality varies between sensors results in probabilistic neighborhoods of different sizes. While it is possible to apply our techniques directly to such cases, it is not clear if additional methods are necessary to provide good performance. Another potential extension is the use of non-metric and asymmetric distances between sensors. For instance, in the case of ocean currents, it may be easier to travel in one direction than another. Such cases have been examined for the classical TSP [34], but not in the case of neighborhoods.

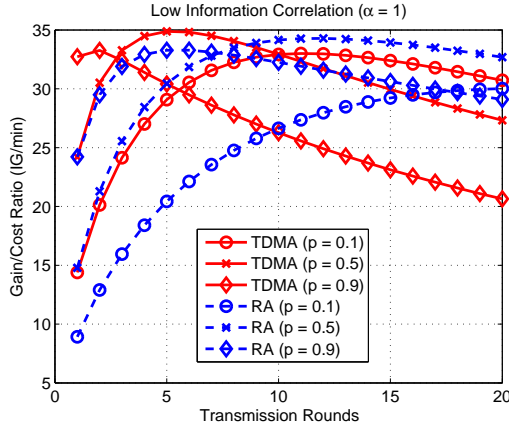
Another avenue to examine is the effect of communication quality dependencies between sensors on the appropriate scheduling parameters and path planning algorithms.



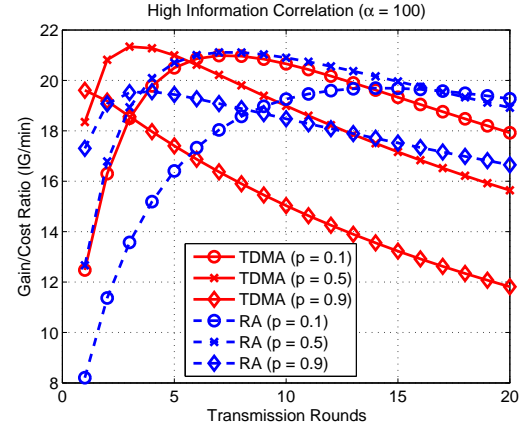
(a) Information gain increases with transmission rounds and asymptotically approaches its maximum



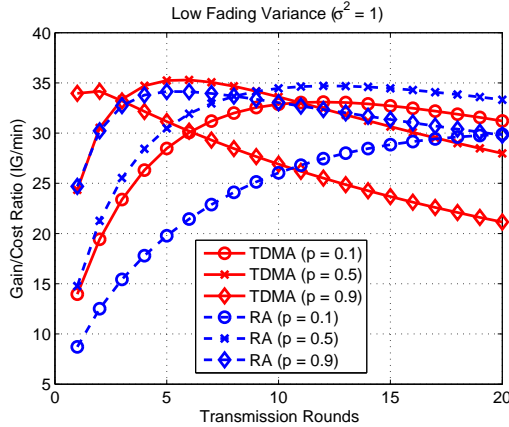
(b) Mission time increases linearly with transmission rounds as the time cost of communication grows



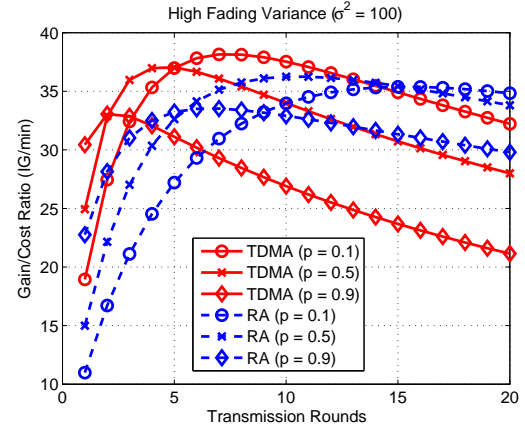
(c) Low information correlation: maximum gain/cost ratio with TDMA at  $p = 0.5$  and five transmission rounds



(d) High information correlation: maximum gain/cost ratio with TDMA at  $p = 0.5$  and three transmission rounds



(e) Low fading variance: maximum gain/cost ratio with TDMA at  $p = 0.5$  and five transmission rounds



(f) High fading variance: maximum gain/cost ratio with TDMA at  $p = 0.1$  (larger neighborhood size) and seven transmission rounds

Fig. 8. Communication-constrained data collection simulations in a  $5 \text{ km} \times 5 \text{ km}$  area with 100 randomly placed nodes. Increasing the value of  $p$  decreases the neighborhood size in the path planning algorithm. A value of  $p = 0.1$  corresponds to approximately 4–5 nodes per neighborhood,  $p = 0.5$  to approximately 3–4 nodes per neighborhood, and  $p = 0.9$  to approximately 2–3 nodes per neighborhood. The maximum gain/cost ratio occurs at different parameter settings depending on the information correlation and gain variance.



Similarly, correlated sensor information could be utilized in the path planning algorithms, rather than considered after planning by the scheduling protocol. Exploiting the locality and submodularity of the information functions [23] may allow for path planning algorithms with performance guarantees to be extended to these cases. Assessing the impact of more sophisticated network structures and cooperation between the sensor nodes is also of interest [35], [36]. Ultimately, better understanding of the connection between path planning and communication has the potential to provide improved performance for robotic data collection.

#### ACKNOWLEDGMENT

The authors gratefully acknowledge Jonathan Binney, Arvind Pereira, Hordur Heidarrson, and Srinivas Yerramalli at the University of Southern California for their insightful comments. Thanks also to Chris Goldfinger of Oregon State University, the captain and crew of the R/V Pacific Storm, and Elizabeth Clarke of the NOAA Northwest Fisheries Science Center for their support of this work.

#### REFERENCES

- [1] R. N. Smith, Y. Chao, P. P. Li, D. A. Caron, B. H. Jones, and G. S. Sukhatme, "Planning and implementing trajectories for autonomous underwater vehicles to track evolving ocean processes based on predictions from a regional ocean model," *Int. J. Robotics Research*, vol. 29, no. 12, pp. 1475–1497, 2010.
- [2] S. L. Nooner and W. W. Chadwick Jr., "Volcanic inflation measured in the caldera of axial seamount: Implications for magma supply and future eruptions," *Geochemistry Geophysics Geosystems*, vol. 10, no. 2, 2009.
- [3] G. Hollinger, S. Yerramalli, S. Singh, U. Mitra, and G. S. Sukhatme, "Distributed coordination and data fusion for underwater search," in *Proc. IEEE Conf. Robotics and Automation*, 2011, pp. 349–355.
- [4] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.
- [5] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*. Princeton Univ. Press, 2006.
- [6] G. Hollinger, U. Mitra, and G. Sukhatme, "Autonomous data collection from underwater sensor networks using acoustic communication," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2011, pp. 3564–3570.
- [7] E. Balas, "The Prize Collecting Traveling Salesman Problem," *Networks*, vol. 19, pp. 621–636, 1989.
- [8] G. Ausiello, V. Bonifaci, S. Leonardi, and A. Marchetti-Spaccamela, "Prize-Collecting Traveling Salesman and related problems," in *Handbook of Approximation Algorithms and Metaheuristics*, T. F. Gonzalez, Ed. CRC Press, 2007.
- [9] G. Ausiello, V. Bonifaci, and L. Laura, "The Online Prize-Collecting Traveling Salesman Problem," *Information Processing Letters*, vol. 107, no. 6, pp. 199–204, 2008.
- [10] M. Goemans and D. P. Williamson, "A general approximation technique for constrained forest problems," *SIAM J. Computing*, vol. 24, no. 2, pp. 296–317, 1995.
- [11] A. Archer, M. H. Bateni, M. T. Hajiaghayi, and H. Karloff, "Improved approximation algorithms for Prize-Collecting Steiner Tree and TSP," in *Proc. IEEE Symp. Foundations of Computer Science*, 2009, pp. 427–436.
- [12] A. Dumitrescu and J. Mitchell, "Approximation algorithms for TSP with neighborhoods in the plane," *J. Algorithms*, vol. 48, no. 1, pp. 135–159, 2003.
- [13] M. de Berga, J. Gudmundsson, M. J. Katz, C. Levcopoulos, M. H. Overmars, and A. F. van der Stappen, "TSP with neighborhoods of varying size," *J. Algorithms*, vol. 57, no. 1, pp. 22–36, 2005.
- [14] D. Bhaduria, O. Tekdas, and V. Isler, "Robotic data mules for collecting data over sparse sensor fields," *Journal of Field Robotics*, vol. 28, no. 3, pp. 388–404, 2011.
- [15] F. Arrichiello, D. N. Liu, S. Yerramalli, A. Pereira, J. Das, U. Mitra, and G. S. Sukhatme, "Effects of underwater communication constraints on the control of marine robot teams," in *Proc. Int. Conf. Robot Communication and Coordination*, 2009, pp. 1–8.
- [16] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, "Data collection, storage, and retrieval with an underwater sensor network," in *Proc. Int. Conf. Embedded Networked Sensor Systems*, 2005, pp. 154–165.
- [17] K. Turgay and C. Erdal, "A MAC protocol for tactical underwater surveillance networks," in *Military Communications Conf.*, 2006, pp. 1–7.
- [18] M. Molins and M. Stojanovic, "Slotted FAMA: a MAC protocol for underwater acoustic networks," in *Proc. IEEE OCEANS Conf.*, 2006, pp. 1–7.
- [19] B. Peleato and M. Stojanovic, "Distance aware collision avoidance protocol for ad-hoc underwater acoustic sensor networks," *IEEE Communications Letters*, vol. 11, no. 12, pp. 1025–1027, 2007.
- [20] A. Goldsmith, *Wireless Communications*. Cambridge Univ. Press, 2005.
- [21] Z. Zvonar and M. Stojanovic, "Performance of antenna diversity multiuser receiver in cdma channels with imperfect fading estimation," *Int. J. Wireless Personal Communications*, vol. 3, no. 1–2, pp. 91–110, 1996.
- [22] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Proc. Uncertainty in Artificial Intelligence*, 2005.
- [23] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *Proc. Information Processing in Sensor Networks*, 2006, pp. 2–10.
- [24] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Optimization*. Wiley, 1994.
- [25] L. Berkhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics*. Springer, 1982.
- [26] H. Singh, A. Can, R. Eustice, S. Lerner, N. McPhee, O. Pizarro, and C. Roman, "Seabed AUV offers new platform for high-resolution imaging," *EOS Trans. AGU*, vol. 85, no. 31, 2004.
- [27] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball, "The WHOI micro-modem: An acoustic communications and navigation system for multiple platforms," in *Proc. IEEE Oceans Conf.*, 2005, pp. 1086–1092.
- [28] P. Qarabaqi and M. Stojanovic, "Modeling the large scale transmission loss in underwater acoustic channels," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*. IEEE, 2011, pp. 445–452.
- [29] —, "Adaptive power control for underwater acoustic channels," in *Proc. IEEE Oceans Conf.*, 2011.
- [30] J. Heidemann, Y. Li, A. Syed, J. Wills, and W. Ye, "Underwater sensor networking: Research challenges and potential applications," *USC/ISI Technical Report ISI-TR-2005-603*, 2005.
- [31] A. Harris, M. Stojanovic, and M. Zorzi, "Idle-time energy savings through wake-up modes in underwater acoustic networks," *Elsevier Journal on Ad Hoc Networks*, vol. 7, p. 770777, 2009.
- [32] A. Molisch, *Wireless communications*. Wiley, 2011.
- [33] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [34] H. Kaplan, M. Lewenstein, N. Shafir, and M. Sviridenko, "Approximation algorithms for asymmetric TSP by decomposing directed regular multidigraphs," *J. ACM*, vol. 52, no. 4, pp. 602–626, 2005.
- [35] Z. Wenyi, M. Stojanovic, and U. Mitra, "Analysis of a linear multihop underwater acoustic network," *IEEE J. Oceanic Engineering*, vol. 35, no. 4, pp. 961–970, 2010.
- [36] M. Vajapeyam, S. Vedantam, U. Mitra, J. C. Preisig, and M. Stojanovic, "Distributed space and time cooperative schemes for underwater acoustic communications," *IEEE J. Oceanic Engineering*, vol. 33, no. 4, pp. 489–501, 2008.



**Geoffrey A. Hollinger** is a Postdoctoral Research Associate in the Viterbi School of Engineering at the University of Southern California. He is currently interested in adaptive sensing and distributed coordination for robots operating with limited communication. Dr. Hollinger's past research includes multi-robot search at Carnegie Mellon University, personal robotics at Intel Research Pittsburgh, active estimation at the University of Pennsylvania's GRASP Laboratory, and miniature inspection robots for the Space Shuttle at NASA's Marshall Space Flight Center. He received his Ph.D. (2010) and M.S. (2007) in Robotics from Carnegie Mellon University and his B.S. in General Engineering along with his B.A. in Philosophy from Swarthmore College (2005).



**Sunav Choudhary** was born in 1988 in Burdwan, India. He received the B.Tech. degree in electronics and electrical communication from the Indian Institute of Technology, Kharagpur, India, in 2010. He is currently a Ph.D. student in the Communication Sciences Institute of the University of Southern California, Los Angeles and is a recipient of the Annenberg Graduate Fellowship. His present research interests are in the field of sparse signal approximation and its applications to underwater acoustic communications.



**Parastoo Qarabaqi** (S'07) received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2007. She is currently pursuing her studies toward a Ph.D. degree in electrical and computer engineering at Northeastern University, Boston, MA. Her research interests include wireless communications, statistical and adaptive signal processing and their applications to underwater acoustic communications.



**Chris Murphy** is a Software Engineer at Bluefin Robotics, and a Visiting Scholar at the Franklin W. Olin College of Engineering. He received a B.S. degree from the Franklin W. Olin College of Engineering, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology / Woods Hole Oceanographic Institution Joint Program. Dr. Murphy's research interests include applied research in marine robotics, particularly compression, networking, and reliable command of autonomous vehicles.

**Urbashi Mitra** received the B.S. and the M.S. degrees from the University of California at Berkeley in 1987 and 1989 respectively, both in Electrical Engineering and Computer Science. From 1989 until 1990 she worked as a Member of Technical Staff at Bellcore in Red Bank, NJ. In 1994, she received her Ph.D. from Princeton University in Electrical Engineering. From 1994 to 2000, Dr. Mitra was a member of the faculty of the Department of Electrical Engineering at The Ohio State University, Columbus, Ohio. In 2001, she joined the Department of Electrical Engineering at the University of Southern California, Los Angeles, where she is currently a Professor. Dr. Mitra has been an Associate Editor for the following IEEE publications: Transactions on Information Theory (2007-2011), Journal of Oceanic Engineering (2006-2011), and Transactions on Communications (1996-2001).



**Gaurav S. Sukhatme** is a Professor of Computer Science and Electrical Engineering at the University of Southern California (USC). He is the co-director of the USC Robotics Research Laboratory and the director of the USC Robotic Embedded Systems Laboratory which he founded in 2000. His research interests are in robot networks, robot perception and on-body networks. He has published extensively in these and related areas. Sukhatme has served as PI on numerous NSF, ONR, DARPA and NASA grants. He is currently PI on an ONR MURI and a Co-PI on the Center for Embedded Networked Sensing (CENS), an NSF Science and Technology Center. He is a fellow of the IEEE and a recipient of the NSF CAREER award and the Okawa foundation research award. He is one of the founders of the Robotics: Science and Systems conference. He is the Editor-in-Chief of Autonomous Robots and has served as Associate Editor of the IEEE Transactions on Robotics and Automation, the IEEE Transactions on Mobile Computing, and on the editorial board of IEEE Pervasive Computing.

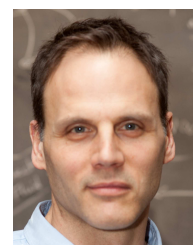


**Milica Stojanovic** (Sm'08, F'10) graduated from the University of Belgrade, Serbia, in 1988, and received the M.S. and Ph.D. degrees in electrical engineering from Northeastern University, Boston, MA, in 1991 and 1993. After a number of years with the Massachusetts Institute of Technology, where she was a Principal Scientist, she joined the faculty of Electrical and Computer Engineering Department at Northeastern University in 2008. She is also a Guest Investigator at the Woods Hole Oceanographic Institution, and a Visiting Scientist at MIT. Her research interests include digital communications theory, statistical signal processing and wireless networks, and their applications to underwater acoustic communication systems. Milica is an Associate Editor for the IEEE Journal of Oceanic Engineering and the IEEE Transactions on Signal Processing.



underwater vehicles.

**Hanumant Singh** received the B.S. degree as a distinguished graduate in computer science and electrical engineering from George Mason University, Fairfax, VA, in 1989 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, MA / Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA, joint program in 1995. He has been a member of the staff at WHOI since 1995, where his research interests include high-resolution imaging underwater and issues associated with docking, navigation, and the architecture of



**Franz Hover** is Finmeccanica Career Development Professor of Engineering, at the Massachusetts Institute of Technology's Department of Mechanical Engineering and the Center for Ocean Engineering. He received the B.S. from Ohio Northern University and the M.S. and Sc.D. degrees from the Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program. Professor Hover has worked broadly in fluid mechanics and marine robotics; his current research centers on design methodologies for complex ocean systems, including power and autonomy.